Chapter 3
Personality in Nonhuman Primates: What Can We Learn from Human Personality Psychology?

Jana Uher

Abstract Primate personality research encounters a number of puzzling methodological challenges. Individuals are unique and comparable at the same time. They are characterized by relatively stable individual-specific behavioral patterns that often show only moderate consistency across situations. Personality is assumed to be temporally stable, yet equally incorporates long-term change and development. These are all déjà vus from human personality psychology. In this chapter, I present classical theories of personality psychology and discuss their suitability for nonhuman species. Using examples from nonhuman primates, I explain basic theoretical concepts, methodological approaches, and methods of measurement of empirical personality research. I place special emphasis on theoretical concepts and methodologies for comparisons of personality variation among populations, such as among species.

3.1 Introduction

All species consist of individuals. These individuals share many characteristics in genome, morphology, physiology, biochemistry, and behavior that define their species membership. But despite this essential similarity, individuals are in no sense uniform; beyond age and sex differences, individuals also differ in their specific genotypic and phenotypic characteristics. The behavioral phenotypes of individuals and their variation within populations are covered by theoretical concepts of personality differences (Stern 1911; Uher 2008a, 2011).

J. Uher (✉)
Comparative Differential and Personality Psychology, Freie Universität Berlin, Habelschwerdter Allee 45, Berlin 14195, Germany
e-mail: uher@primate-personality.net

A. Weiss et al. (eds.), *Personality and Temperament in Nonhuman Primates*, Developments in Primatology: Progress and Prospects, DOI 10.1007/978-1-4614-0176-6_3,
© Springer Science+Business Media, LLC 2011

The scientific study of personality differences in human (Galton 1869; Stern 1911; Allport 1937) and nonhuman primates (Crawford 1938; Yerkes 1939; Hebb 1949) started about 100 years ago. Whereas human personality research has evolved into a discipline of its own within psychology, nonhuman primate personality research developed only incompletely within several heterogeneous disciplines. Yet many challenges are structurally similar if not identical to those in human research, allowing primate researchers to profit greatly from the theoretical, methodological, and statistical advances made in psychology.

In this chapter, I introduce theoretical concepts, methodological approaches, and methods of measurement from human personality psychology, and discuss their suitability for empirical studies in nonhuman primates. My special concern is to show how established theoretical concepts and methodologies for cross-cultural comparisons of human personality variation generalize to cross-population comparisons of personality variation within and across species. Using examples from nonhuman primates, I discuss theoretical foundations and typical methodological challenges that provide the necessary background for empirical research. How can we compare individuals when they are all unique? What role do situations play in studying individuals? How can personality variation be compared among different species? How can we decide what is important to study within a species? And what methods can we use to measure the personality of nonhuman primate individuals? This chapter explores theoretical concepts and suitable methodological tools for these and other puzzling issues in empirical research on primate personality.

3.2 Theoretical Concepts for Primate Personality Research

To explain personality differences in human primates, psychologists have developed various classical schools of thinking. They differ in basic ideas of man, theoretical concepts, investigative methods, and explanatory approaches (Buss 1991; Funder 2007; Cervone and Pervin 2008). Perhaps most commonly known outside psychology are the psychoanalytic approaches grounded in Freud's theories that assume infantile psychodynamics determine an individual's personality. Humanistic psychologists try to explain the individual through its unique conscious experience of the world driven by its free will and the striving for personal growth and for an understanding of the meaning of life.

To oppose the mainly introspective methods of these approaches, which hinder empirical investigation, behaviorists tried to explain an individual's personality as a result of its learning history. They assumed individuals are born as *tabulae rasae*, as "blank slates" with no innate content, whose development is largely determined by acquired stimulus-response connections. Cognitive psychologists filled the behaviorists' black box with structures of information processing that explain personality differences with variations in the architecture and processing parameters of the individuals' cognitive systems. Social constructivist psychologists view personality as created through interactions and negotiations with others. Developmental psychologists focus on

continuous dynamic transactions between developing individuals and their changing environments to investigate processes of individual personality development.

In search of the biological basis of personality, biological psychologists and neuropsychologists study processes in the neural, hormonal, and immune systems that underlie observable individual differences. Behavior genetic approaches estimate average contributions of genes and environment to behavioral differences on the basis of twin and adoption studies. They are increasingly refined by molecular genetic approaches that study the transaction between specific genes and specific environments over the course of life. A promising approach models systematic transactions between intertwined genetic and environmental influences (Johnson 2007). Finally, evolutionary psychologists understand personality differences as proximate mechanisms that have evolved in adaptation to environmental conditions.

All these classical schools of thought with their different philosophical, theoretical, and methodological principles have contributed to our understanding of human personality. Clearly, some of them, such as psychoanalytic or humanistic schools that rely on introspective methods, are not suitable for empirical research in nonhuman primates. Behaviorism was too one-sided because it neglected genetic influences. Yet many others have broad intersections with nonhuman research, in particular those focusing on information-processing, genetics, neurobiology, ontogeny, and evolution. They try to unravel mechanisms and processes governing observable behavioral differences – provided we already have a sketchy road map of what kind of individual differences a species exhibits. It was probably not by chance that one of the oldest and most influential schools in personality psychology, trait psychology, focuses on this essential first task of measuring and cataloguing individual differences. Stern (1911) laid the methodological foundations of empirical and statistical approaches that form the basis of much of today's personality psychology. They also provide an excellent foundation for empirical research on nonhuman personality.

3.2.1 Variable-oriented and Individual-oriented Perspectives on Individuals

Primate individuals exhibit individual-specific behavioral patterns that are commonly construed as their personality. Individual-specificity implies that these patterns are relatively stable within each individual over time, and that the individuals vary in the degree to which they exhibit certain behavioral patterns (Uher 2011). Their empirical interindividual variation across the composite of the population can be described with theoretical dimensions of personality differences (Stern 1911). Theoretical conceptions of such behavioral patterns are also called personality traits, personality constructs, or trait constructs; accordingly, the dimensions that describe their interindividual variation are also called trait dimensions or personality dimensions.

In explanatory models of personality, individual-specific patterns of behavior are interpreted as reflecting the individuals' psychobiological organization that determines their unique adaptations to their environments (Allport 1937). Personality traits are

thus conceived as reflecting behavior-regulating mechanisms that can have genetic, physiological, cognitive, motivational, and behavioral components (Buss 1991; Mischel et al. 2003; Funder 2007).

To make these behavioral patterns accessible for empirical research, Stern (1911) introduced the differential perspective to psychology that was the groundbreaking shift in viewpoint from the average individual to differences among individuals. He laid the methodological foundations of empirical personality research by conceiving two complementary methodological perspectives.

The first perspective focuses on the measurement variables. Variable-oriented analyses address the individuals' relative positions along shared trait dimensions. First, they analyze the statistical distributions of trait scores in specified populations. In many populations, many trait scores, such as human extraversion, are normally distributed. Most individuals' scores center around the mean of the dimension, and only a few individuals are on its extremes. If a trait's variability is limited on one side of the dimension, the distribution pattern can be skewed. On aggressiveness, for example, most humans score rather low, and only a few are high scoring. Furthermore, variable-oriented analyses address the covariation of individual trait score distributions among various trait dimensions in a population which I explain further below in the section about personality taxonomy. That is, variable-oriented analyses characterize the population.The second perspective focuses on individuals. Individual-oriented analyses address the individual's unique configuration of its relative positions across multiple trait dimensions that have been identified in its population with variable-oriented analyses. This allows us to quantify the individuals' uniqueness based on their empirical comparability along shared dimensions. Quantification of an individual's personality thus depends on the personality variation of the other individuals to which it is compared and that are called the reference population.

Individual-oriented analyses rely mostly on standardized scores that depict the individual's relative scores in comparison to those of other individuals in its sample. Absolute score profiles, in contrast, are confounded with the mean profile of the sample. For example, since all individuals generally score higher on locomotion than on social play, absolute scores may fail to reveal that some individuals may score higher than others in social play and lower than others in locomotion (as in Suomi et al. 1996). The pattern of an individual's relative trait scores can be illustrated as a profile across trait dimensions; the shape of this trait profile characterizes the individual (Stern 1911; Cairns et al. 1998; Mervielde and Asendorpf 2000).

Standardized personality profiles can be illustrated with behavioral data from great apes. In a methodological study, Uher et al. (2008) repeatedly observed 20 great apes (five each of bonobos, chimpanzees, gorillas, and orangutans) in 14 different laboratory test situations and two different group situations. They studied 19 different personality trait constructs that they measured with 76 behavior variables, most of which could be obtained from all four species. The data were analyzed systematically from both variable-oriented and individual-oriented perspectives.

Figure 3.1 shows z-scored trait profiles from two individuals in that study. A z-score is a measure of deviation from the sample's mean that is standardized

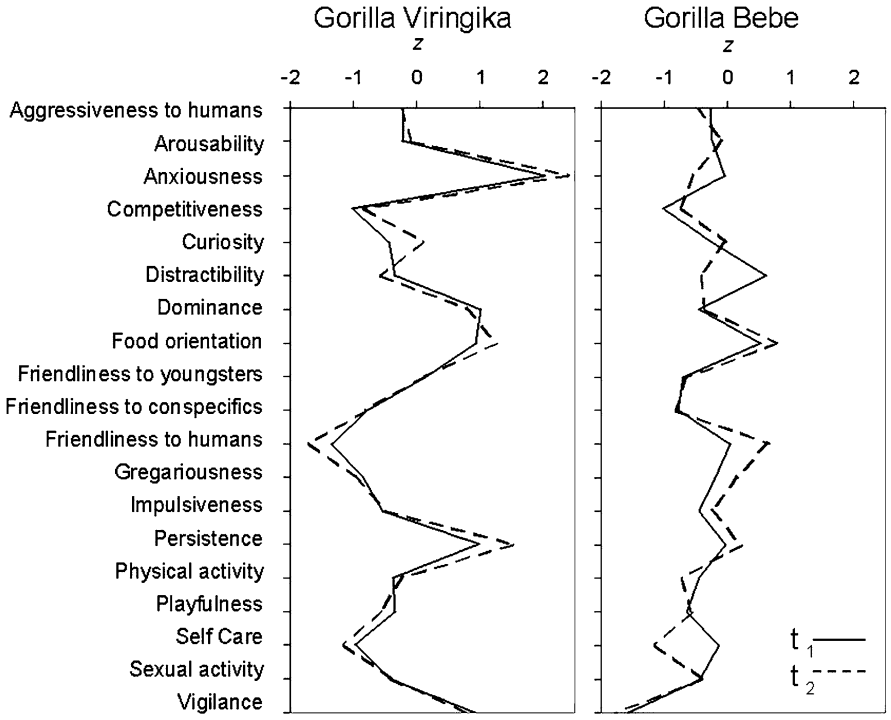

Fig. 3.1 Personality profiles of two individuals based on ethological measures of behavior obtained in a series of 14 laboratory tests and group observations in two different group situations. All trait scores represent behavioral measures that were aggregated over several occasions of measurement. For details on ethological behavior measurement see Uher et al. (2008). The *z*-standardized trait scores depict the individuals' positions on each trait dimension in relation to those of the other individuals of the sample. The sample's mean score is thereby 0, and the standard deviation is 1. The data were aggregated and the aggregate scores were standardized across individuals separately within two nonoverlapping test periods; t_1 is the first test period, t_2 second is the test period 3–6 weeks later

such that the sample mean is 0, and the sample standard deviation is 1. Standardization allows three kinds of direct comparisons: (1) an individual's score can be placed within the trait distribution of the population, (2) its scores can be compared across different traits, and (3) different individuals can be compared on the same trait.

Viringika, for example, is high scoring on anxiousness; her trait score is about two standard deviations above the sample mean. Her scores on dominance, food orientation, and persistence are also about one standard deviation above average. But her scores on competitiveness, friendliness to humans, and self-care are one standard deviation below average. Across traits, these deviations are the most pronounced in her profile, whereas her scores on other traits are rather average. Comparing Viringika's trait profile with that of Bebe, one can see that both are more food oriented than the sample average, but that Viringika is even more "greedy" than Bebe. These females score equally low on competitiveness.

Their scores on dominance, in contrast, are quite different; Bebe tends to be submissive, whereas Viringika is quite dominant. How do we know that these behavioral scores can be used to infer these individuals' personality?

3.2.2 *Temporal Consistency of Interindividual Behavioral Differences*

The everyday connotation of the word "trait" already implies characterization by lasting attributes. Personality traits imply characterization by relatively stable individual-specific behavioral tendencies. The stability criterion is important since interindividual variation can also derive from momentary behavioral fluctuations that are unrelated to the individuals' lasting behavioral tendencies. Thus, measuring personality differences in the flood of individual behavior requires repeated observations and evidence of temporal consistency. A basic criterion of personality measurement is therefore test–retest reliability. Individuals having low scores on a personality trait should retain their position relative to other individuals in retest assessments at least over intermediate time periods. Variable-oriented test–retest reliability means that the individuals' rank orders on that dimension should correlate over time. Individual-oriented test–retest reliability means that the individuals should retain their individual-specific behavioral patterns; their individual profile shapes across multiple trait dimensions should correlate over time (Cairns et al. 1998).

The fluctuating nature of behavior often hinders establishment of test–retest reliability and entails particular methodological difficulties (Hebb 1949; Stevenson-Hinde et al. 1980; Suomi et al. 1996). A strategy to reduce the impact of random variation and measurement error, and to increase the reliability of personality measurement is aggregation at least over multiple occasions, if not over different trait-related behaviors and situations (Rushton et al. 1983).

These methodological principles can be demonstrated with results from Uher et al. (2008). In that study, the individuals were observed repeatedly in the same test and group situations over a period of about 2–3 weeks. After a break of about a fortnight, all individuals were again observed repeatedly in the same behaviors and situations in a second 2- to 3-week period. Overall, each individual was observed for more than 67 h within a 50-day period. Due to the intense and repeated observations in this design, the behavioral raw data could be aggregated within each of the two nonoverlapping periods and analyzed for temporal reliability between them. Mean variable-oriented temporal reliability of the 76 behavioral variables was high ($r = 0.78$) as was temporal reliability of 19 trait indices each composed of several different trait-related behavioral variables ($r = 0.77$). This shows that the data were sufficiently aggregated and that reliable personality measures were obtained. Personality differences can thus be measured in great ape behavior as reliably as those in human behavior – provided the data are aggregated sufficiently (Uher et al. 2008).

Insufficient behavior observation can result in unreliable personality measures that compromise comparisons and coherence with measures obtained by other

methods, such as ratings. For example, focal samples of 15 min are obviously insufficient to measure personality in chimpanzees reliably outside controlled laboratory settings. Comparisons of behavior observations across such extremely short time periods yielded only low to zero reliability scores (Vazire et al. 2007) that are best interpreted as reflecting estimates of the daily fluctuations of behavior rather than of the individuals' personality. Comparisons of such unreliable measures with rating measures based on the raters' mental aggregations of everyday observations over 7 years (as in Vazire et al. 2007) are therefore necessarily compromised. Based on these unreliable measures it cannot be concluded that ethological behavior measures would be per se unreliable or even inferior measures of personality (as assumed by Vazire et al. 2007; Gosling 2008).

Reliable measures of personality, whether ethological behavior measures or rating measures, can only be obtained with sufficient aggregation across repeated observations. When this principle is considered adequately, behavioral personality measures were shown to be as reliable as those obtained with ratings in nonhuman primates (Uher and Asendorpf 2008). Similarly, raters must have sufficient observational experiences with the target individuals; ratings provided by raters who hardly know the individuals will be meaningless.

Most primate studies focus on variable-oriented analyses. Yet, temporal stability at the population level may mask changes occurring at the individual level. Even high rank-order stability does not mean that each individual retains the same relative position over time. Instead, a few individuals may change, while the majority of individuals remain the same. Such differences in individual trajectories are essentially a question of stability, gradual change, and long-term development of individual personality that can only be studied with individual-oriented analyses. These analyses can reveal information beyond those shown in variable-oriented analyses and can therefore contribute meaningfully to personality studies (Block 1971; Magnusson 1988; Mervielde and Asendorpf 2000). To show more of their potential, I will discuss individual-oriented analyses often in this chapter.

Figure 3.1 illustrates the test–retest reliability of individual trait profiles. Profiles indicated by continuous lines are based on aggregated measures obtained in the first observation period of the Uher et al. (2008) study; those indicated by broken lines derive from the second period 3–6 weeks later. Given that these profiles were measured and standardized independently in two nonoverlapping observation periods, their shapes are remarkably similar. These findings show that these behavioral profiles are reliable measures of the individuals' personality. Among individuals, test–retest reliability scores varied from $r = 0.49$ to 0.94 for profiles across all 76 single behavioral variables and from $r = 0.38$ to 0.97 for profiles across 19 composed trait indices. Temporal correlations were significant ($p < 0.05$) in all single behavior profiles, and in 90% of the composed trait profiles. This shows that stability and change are also manifested at the individual level. It suggests that some individuals are more consistent in their behavior, whereas others may be guided more by environmental influences. Hence, behavioral consistency itself seems to be interindividually different (Caspi and Roberts 1999; Funder 2007).

3.2.3 The Role of Situations in Personality Research

Early personality theorists had assumed not only temporal consistency but also substantial consistency across situations to be central to personality. When Hartshorne and May (1928) reported low consistency in the behavior of 850 school children across different situations, the concept of personality seemed to be challenged fundamentally. It culminated in Mischel's (1968) finding that cross-situational consistency in behavior rarely exceeds the "magic" correlation of $r = 0.30$. Individual behavior appeared to be highly situation specific rather than individual specific and cross-situationally consistent. These puzzling findings provoked the person-situation controversy that lasted four decades in psychology (Mischel 1968; Funder and Colvin 1991; Fleeson 2004; Funder 2006).

When looking for consistency across situations, primate researchers came across exactly the same findings. In their famous series of studies on personality differences in rhesus macaques (*Macaca mulatta*), Stevenson-Hinde et al. (1980) reported that reliable behavior measures were lacking significant correlations across situations. But, instead of reflecting a "failure to look at appropriate measures rather than a characteristic of the … [individuals] themselves" (p 508), these findings mirror the core issues of cross-situational consistency. They show that careful methodological considerations are needed to avoid misinterpretations of empirical findings.

First, cross-situational consistency is no mere illusion. Moderate correlations show that individuals do display some consistency in their behavior across situations; it is just less than initially expected, and behavioral correlations across situations are lower than their correlations over time. This also shows that situations exert significant impacts on individual behavior because individuals respond to them differently (Mischel and Peake 1982). Individual-oriented analyses can reveal whether such differences are individual specific; they quantify and illustrate the individuals' unique patterns of responsiveness to different situations in behavior profiles across situations. Such a situation-behavior profile depicts the individual's scores on the same trait dimension measured in different situations (Mischel et al. 2002).

If individuals' trait scores are standardized within each situation, the profile informs about situational influences that are specific to the target individual. For example, most chimpanzees react more fearfully to snakes than to petrol cans (Goodall 1986). Unless their responses are standardized within situations, most chimpanzees will therefore exhibit higher fear scores for snakes than for petrol cans. Their individual situation-behavior profiles would be confounded with that of the average chimpanzee. After standardization, chimpanzees that are generally more fearful toward everything will have positive z-scores. Those individuals that are less snake-fearful than the average chimpanzee will have negative z-scores for snakes. Independent of that, some chimpanzees will show large positive z-scores for snakes as compared to their z-scores for other situations; these chimpanzees are more specifically snake-fearful.

Such differences in situational responsiveness are individual-specific if they are temporally consistent (Mischel and Shoda 1995). Individual-oriented test–retest reliability analysis of situation-behavior profiles is illustrated in the Uher et al. (2008) study that obtained behavioral data for the same traits in various situations. Aggressiveness, for example, was measured in four laboratory-based test situations involving familiar keepers, observers entering neighboring cages, friendly masked humans offering food, and playbacks of radio news records. The average correlation of aggressiveness scores across these situations on the sample level was $r = 0.25$; yet, the individuals' aggressiveness-situation profiles correlated on average $r = 0.77$ over time (3–6 weeks), ranging from an outlier with an almost inversed profile of $r = -0.49$ to 0.99 ($N = 16$; Uher et al. 2008). This means that the individuals differed substantially in how strongly they responded with aggression to these four situations, yet within each individual, aggressiveness patterns across these situations were fairly stable.

Test–retest reliable situation-behavior profiles reflect consistent interactional patterns between situations and the individuals' responses over time. Individuals may not only respond to situations in particular ways, they may also actively choose particular environments that are suited to their personality; they may evoke certain reactions from the environment, in particular, from their social environment; and they may also actively shape their environments. Such interactions may be the mechanisms behind the increasing matches between certain personalities and certain environments, and thus behind continuity in personality development (Magnusson 1988; Matthews et al. 2003).

Situations, conceived as complex constellations of stimuli, vary in how they permit personality differences to emerge. Two qualitatively different aspects, situational strength and trait-relevance, are distinguished. Situational strength denotes how compelling a situation is for the individuals' behavior. For example, variations in aggressiveness might emerge most clearly in situations that typically elicit low to moderate aggression. Situations that permit easy emergence of personality differences are referred to as weak situations (Mischel 1977). Strong situations, by contrast, may mask interindividual variability because they force behavior into specific channels by either inhibiting the behavior substantially or by evoking heightened responses from all individuals (Tett and Guterman 2000). For example, most captive primates react strongly to veterinarians with blow guns in front of their cages, making interindividual differences less pronounced.

The second aspect of situations is trait relevance, which refers to the type of information to which the individuals are responding. That a behavior cannot be observed does not necessarily mean that the individual has a low trait score, or that the assumed trait construct is a mere theoretical hypothesis without any empirical relations to observable behavior. Situations have to activate relevant behavior. For example, aggressions are responses to stimuli indicating that aggressive behavior might be functional. Individuals that are more sensitive to them and that react more quickly or more intensely with aggression than others are assumed to be more aggressive (Tett and Guterman 2000; Capitanio 2004).

3.2.4 Individual Response Specificity

Typically, trait constructs are inferred from different behavioral responses. For example, human shyness is inferred from long pauses in speech, hesitant speaking, gaze aversion, or restricted gestures (Asendorpf 1988). Chimpanzee arousability in prefeeding contexts can be inferred from rocking, grinning, vocalizing, or pacing (Uher et al. 2008). Since these responses are assumed to indicate the same trait, they should be correlated. Surprisingly, they are not; both studies report low to zero correlations among the different behavioral indicators of these two personality traits.

We can gain some understanding of this puzzling finding with individual-oriented analyses. Analogous to cross-situational consistency, correlations among behavioral trait indicators can be low on the sample level because they lack validity for the trait in question, or just vary randomly. Yet they can also be low due to stable individual response specificity. It imposes methodological difficulties since restricting personality measurement to single behaviors can result in misclassifying those individuals who primarily exhibit behaviors that are not measured. In fact, traits can often be inferred from a variety of responses that are not necessarily shown by all individuals (Asendorpf 1988; Marwitz and Stemmler 1998; Uher et al. 2008).

Individual response specificity can be analyzed and illustrated in individual response profiles that depict their scores across different behavioral indicators of the same trait. Behavior measures are standardized within the sample because absolute behavior scores would confound interindividual differences with sample-level differences. For example, while all chimpanzees may generally show more rocking than pacing in prefeeding situations, some individuals may show, in comparison to others, more pacing than rocking. Standardized behavior scores thus inform about individual-specific patterns of behavioral trait indicators. They also allow comparisons of different types of behavior measures such as durations, latencies, and frequencies that can be neither directly compared nor simply averaged since they may be distributed differently.

To capture such interindividual differences in response specificity, the Uher et al. (2008) study measured most traits with multiple behaviors. Figure 3.2 illustrates individual arousability profiles across different arousal responses (rocking, grinning, vocalizing, or pacing) of four chimpanzees prior to their noon feedings. The *z*-scores indicate the individuals' relative positions on each response variable and allow direct comparisons. One can see, for example, that Frodo showed pleasure grins much more often than the others; he scored three standard deviations above the sample's mean. Robert and Fraukje were rocking much more often than Dorien or Frodo; they scored two standard deviations higher than the others. Their particular profile shapes illustrate the typical arousal responses of these individuals. For Fraukje, it was most characteristic to rock, vocalize, and change position when awaiting the feeding, whereas she hardly ever paced. Comparison of these four response profiles also shows that measuring arousability only with rocking would misclassify Frodo.

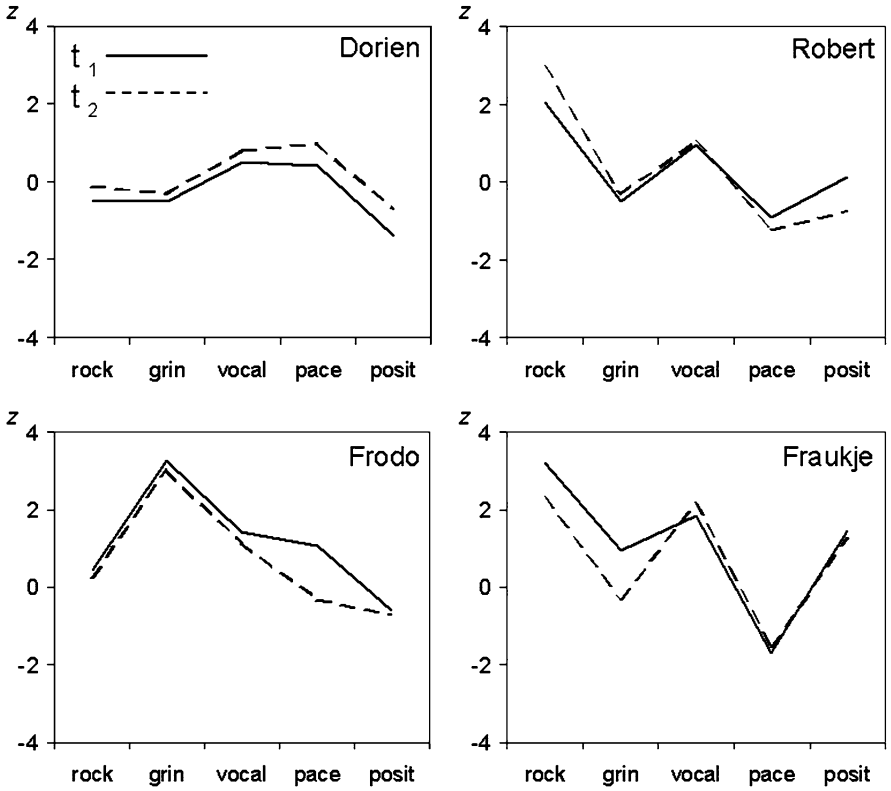

Fig. 3.2 Response profiles depicting individual response specificity. Five arousal-indicating behaviors observed prior to the noon feeding were analyzed: rock = rocking, grin = pleasure grin, vocal = vocalizing, pace = pacing, posit = changing position, defined as rising from the waiting position, and sitting down again or staying within 1.5 m from the original place within 10 s. The *z*-scores depict the individuals' positions on each response in relation to the other individuals of the sample; the sample's mean score is thereby zero. t_1 is the first test period, t_2 is the second, nonoverlapping test period 3–6 weeks later. Within each test period separately, the data were aggregated across multiple occasions of observation; the aggregated scores were then standardized across individuals

Empirical test–retest reliability reveals whether such response profiles are individual-specific. On the sample level, variable-oriented correlations of the individuals' rank orders among the different behavioral indicators of some of the traits studied by Uher et al. (2008) were on average $r = 0.16$. This reflects that individuals can vary in which and how many of multiple trait-related behaviors they show. Yet, on the individual level, individual response profiles consisting of the individuals' relative scores on different trait-related behaviors correlated on average $r = 0.66$ over 3–6 weeks, indicating temporally consistent individual response specificity. Figure 3.2 illustrates these findings. The shapes of the individual response profiles of the

four chimpanzees measured in the first observation period are very similar to those measured in the second nonoverlapping observation period, as indicated with continuous vs. broken lines.

3.2.5 Personality Types

Although individual profiles are distinct and unique, there may be groups of individuals showing similarities in their profile shapes. The response profiles of Fraukje and Robert in Fig. 3.2, for example, have strikingly similar shapes. This may indicate a shared response profile *type*. Similarly, there may be individuals sharing similar situation-behavior profiles or similar personality profiles. Such personality types can be identified statistically with cluster or Q-factor analysis; they represent prototypes of similar individuals (Asendorpf and van Aken 1999).

Extreme scores on a single trait dimension are also sometimes referred to as types, for example, the extravert type. These "univariate" types are special cases of the configurational "multivariate" types that are based on multiple traits. To my knowledge, personality types have not yet been analyzed empirically in behavior-based studies of nonhuman primate personality. Yet some rating-based studies in chimpanzees identified distinct personality types that were defined by characteristic trait score patterns, such as those labeled as "socially confident" (Murray 2002) or "deferent apprehensive" types (Martin 2005).

3.2.6 Personality Taxonomy

Typically, certain personality traits go together in a population. Their covariation can be subsumed statistically within broader, higher order trait constructs underlying this shared covariation, thus making the less complex trait constructs subtrait constructs of the emergent, more complex trait constructs. Such patterns of variable-oriented trait correlations can be analyzed with multilevel, cluster, or factor analysis. They can be organized in hierarchical trait taxonomies. At the top of such hierarchies are a few abstract trait constructs, often called personality factors, which summarize the shared variance of the correlated lower order traits they comprise (see King and Weiss 2011). Preferably, linear, relatively independent factors are extracted that do not overlap too strongly in the lower order trait constructs they summarize (Eysenck 1990; Matthews et al. 2003).

The concept of trait hierarchies or trait taxonomies shows that personality factors can explain more diverse behaviors than each of their lower order trait components alone. Thus, factors permit parsimonious and comprehensive descriptions of personality differences. Identifying multiple personality factors increases the possibilities to explain complex observable diversity among individuals. Unique individual configurations of factor scores; that is, individual personality profiles depict unique personalities (Capitanio 2004; Uher 2008a, b, 2011).

3.2.7 Theoretical Concepts for Cross-Species Comparisons

Comparative personality research merges three areas of emphasis: the individuals' uniqueness, their comparability, and their universality. Uniqueness and comparability are studied based on individual- and variable-oriented analyses. Since the individual's relative position on a trait dimension depends on the scores of those individuals to whom it is compared, the differential perspective implies population dependency. If reference populations change, all trait scores very likely change, too. Hence, studies of universality, which address whether particular personality dimension are common to different populations, are based on specifications of the studied reference populations and on methodologies for comparisons of personality variation among populations (Uher 2008a, b).

In humans, reference populations are typically defined by social criteria such as culture, language, or nationality. Quests for human universals thus refer to trait constructs that are applicable to all humans regardless of their cultural, language, or national background. Similarly, populations in nonhuman primates can be defined by their geographical distribution or living environment. For example, the universality of chimpanzee personality trait constructs can be studied by comparing populations living in wildlife sanctuaries with those living in zoological parks (King et al. 2005). When we define reference populations by biological criteria such as breed, subspecies, or species, universality can be studied on more general population levels. Comparing species nested in the biological classification, such as within genera, families, or orders, could show whether some personality constructs can be assumed to describe behavioral variation that is, for example, uniquely macaque, uniquely pan, uniquely hominoid, or uniquely primate (Uher 2008a; King and Weiss 2011).

This suggests that theoretical concepts and methodologies for comparisons of human cultures can be generalized to comparisons of species. Cross-cultural personality research has shown that personality variation as dimensions of stable interindividual behavioral variation can also be conceptualized *across* different populations (Leung and Bond 1989). Generalizations of these concepts yield three basic kinds of personality dimensions (Uher 2008a). Population-specific trait dimensions differentiate individuals of only one particular population, but not those of other populations. Universal trait dimensions, in contrast, differentiate individuals across different populations. This means that individuals of several considered populations differ along these dimensions. Two kinds of universal trait dimensions can be distinguished. Weak universal traits are dimensions on which populations show similar means and variances. Strong universal traits are dimensions on which populations exhibit significant mean level differences. The latter are thus also population-comparative trait dimensions, yet should not be mistaken for behavioral dimensions that differentiate populations without also differentiating individuals. Comparisons of personality differences among populations are ultimately always based on interindividual variability within each population. These basic kinds of traits can be analyzed with population specific, universal, and population-comparative analyses (in particular factor analysis; for details see Uher 2008a, b).

Evolutionary theory provides supportive arguments for the existence of behavioral differences that can be described with these kinds of trait dimensions (see Sih 2011). In evolutionary research, personality differences are understood as behavioral strategy differences based on trade-offs with different costs and benefits that have evolved in patchy and changing environments, and that reduce the pressure of competition among members of a species. Accordingly, population-specific personality traits could reflect behavioral differences that are niche-differentiated adaptations (Tooby and Cosmides 1989). For example, in adaptation to an arboreal life in swampy rainforests where food is difficult to furnish, orangutans could have evolved individual differences in behavior that are not displayed by other primate species, and that can therefore be interpreted as an orangutan-specific personality trait. Phylogenetic hypotheses, in contrast, suggest that some personality traits could reflect interindividual differences in behavior that can be explained as homologies inherited from common ancestors. For example, all living macaque species may show interindividual differences in sociability. That is, individuals within each species differ in their degree of sociability. Observations suggest that at least some macaque species may thereby differ in their average scores, such as bonnet macaques that scored higher on average sociability than pigtailed macaques (Capitanio 2004). This could indicate that sociability is a strong universal trait dimension that differentiates both individuals and species. Such findings may reflect behavioral patterns inherited from a common macaque ancestor that could be illuminative for theories of speciation (Uher 2008a, b).

These three basic kinds of trait dimensions can co-occur in the personality structure of a population. For example, the personality structure of orangutans may comprise some weak and/or strong universal trait dimensions they share with other species as well as some orangutan-specific trait dimensions. Personality variation of populations can thus be compared quantitatively in shared weak and strong universal trait dimensions on which populations may exert different positioning effects. Personality variation can also be compared qualitatively in terms of differences in the populations' hierarchical trait taxonomies, and thus in their personality factors. The results are referred to as the populations' patterning effects (for details see Uher 2008a, b).

Identification of such differences among populations could be informative for theories and models developed to explain the causation, function, adaptation, and phylogeny of individual differences in behavior. Mean level differences among species, for example, could be associated with ecological differences in predation risk or food density thereby indicating possible functionality and adaptivity of individual behavioral differences. Given their uniqueness, species-specific trait or factor variations could be particularly illuminative regarding ecological functions of behavioral variation and processes of speciation (Capitanio 2004; Uher 2008a, b). For example, if the trait construct of conscientiousness could only explain behavioral variation in humans, this could reveal important information about unique antecedents of human evolution. Personality trait dimensions shared by closely related species, in turn, may indicate behavioral strategy differences that could be interpreted as homologs inherited from common ancestors, whereas those shared by distantly related species occupying similar ecological niches could reflect

analogs evolved in adaptation to similar environments (Gosling and Graybeal 2007; Uher 2008b).

The concept of hierarchical trait taxonomies emphasizes that species comparisons depend on comprehensive empirical models of the structure of interindividual behavioral variation of species. For example, if indicators of conscientiousness are not studied in a species, empirical results cannot be interpreted as indicating that this trait construct is not applicable to that species (Weiss et al. 2006). This has strong implications for the validity of species comparisons and may bias inferences on possible antecedents of the emergence of behavioral variation explained with that construct. Methodological approaches are necessary that allow to identify comprehensive and ecologically valid models of the species' hypothetical true trait taxonomies (Uher 2008a, b).

3.3 Methodological Approaches to Primate Personality

Establishing representative and comprehensive taxonomic models of interindividual behavioral differences in a population encounters two crucial bottlenecks: comprehensive selection and systematic reduction. First, all potential trait constructs should be selected comprehensively to avoid ignoring important domains of personality variation in the target population. Second, these trait domains should be analyzed empirically for dimensions of test–retest reliable interindividual behavioral variation that must then be reduced systematically to broad personality factors that summarize their shared variance. Bias or arbitrariness in either of these processes reduces the representativeness of empirically identified hierarchical trait taxonomies, which may compromise inferences on patterning and positioning effects of populations. Whereas reduction procedures are largely based on statistical tools, and thus on statistical criteria, selection procedures require stringent rationales to ensure that a comprehensive pool of potential trait constructs and measures is entered into the identification process (Uher 2008a, b).

3.3.1 A Taxonomy of Methodological Approaches

The diversity of behavioral variations within, and especially across, species makes it difficult to decide what to study. How did human personality psychology solve this problem? To ensure comprehensiveness, some of the founders of trait psychology reasoned that "those individual differences that are most salient and socially relevant in people's lives will eventually become encoded into their language; the more important such a difference, the more likely it is to become expressed as a single word" (John et al. 1988, p 174). Hence, natural language is assumed to be a comprehensive pool of human personality descriptors. This approach provides the basis for much of contemporary research on human personality.

Based on this lexical hypothesis, Allport and Odbert (1936) went through about 550,000 words of the 1925 edition of Webster's New International Dictionary, and identified 17,953 terms describing personality differences. From this enormous list, they further extracted 4,500 adjectives that describe observable and lasting traits. This list set the stage for empirical models of the human personality structure based on different reduction methods. The factor analytic reduction to five broad personality factors (Extraversion, Agreeableness, Conscientiousness, Emotional Stability, and Openness), the so-called Big Five, has received substantial empirical support in various languages (Goldberg 1990; John 1990; de Raad and Barelds 2008).

The lexical hypothesis also implies, however, that the human lexica cannot serve as comprehensive pools of animal personality descriptors. There is no reason to assume that humans have codified in their natural language an equally systematic body of trait-related descriptors for interindividual behavioral differences in other species with which they generally interact little or not at all. The English language, for example, evolved primarily in parts of the world that are outside primate habitat regions. How could English-speaking people have developed a systematic vocabulary that describes all salient and socially relevant behavioral characteristics of nonhuman primate individuals when they do not even encounter such individuals regularly in their daily lives? Over the last decades, primate researchers invested considerable efforts to describe the extraordinary variety of primate behavior in comprehensive ethograms. Rarely are single words, especially adjectives, sufficient to describe and differentiate the complex courses of motion and facial expressions of nonhuman species in a way that all other people, including laypeople, can readily understand their meaning without any further explanation. Human trait descriptors, in contrast, can convey precise information, for example, about specific facial expressions of human emotions.

This does not exclude, however, that some lexical trait descriptors can also be useful to describe interindividual differences in primate behavior, as I will show below. Ultimately, any scientific investigation has to rely on human language. But the usage of lexical trait descriptors in subjective ratings as one out of several methods of personality measurement has to be clearly distinguished from the methodological approach that is used in order to decide at first which behaviors shall be studied for interindividual differences in a species. The lexical approach, which turned out to be enormously productive for human personality psychology, therefore fails as a systematic methodological approach to actualize comprehensive selections of trait constructs and trait measures for nonhuman species in order to empirically establish comprehensive trait hierarchies of their interindividual behavioral variation (Uher 2008b; Uher and Asendorpf 2008).

Besides lexical approaches to human personality, various other methodological approaches are used to decide what to study in human and nonhuman populations; they can be taxonomized into five major groups. (1) Nomination approaches rely on human observers who nominate trait constructs or measures based on their perceptions of the individual behavioral characteristics of the target population and on implicit theories they have developed about it. (2) Adaptive approaches derive trait constructs from ecological and evolutionary theories on interactions of populations with their environments to identify domains of interindividual behavioral variation in response to present and/or past adaptive problems. (3) Bottom-up or emic

approaches use naturally evolved, complex systems inherent to the species such as language (exclusively in humans), behavioral, or neurobiological systems to derive trait constructs and measures. (4) Top-down or etic approaches import trait constructs and measures from other species to look for differences and similarities in their patterning effects. (5) Eclectic approaches capitalize on findings and methodologies of the other approaches without holding to a single approach. Since these approaches were developed for various aims and purposes, their rationales are not necessarily suited to identify ecologically valid and comprehensive models of trait taxonomies (Uher 2008a, b).

Systematic bottom-up/emic approaches, such as that of the lexical approach to human personality, enable comprehensive selections because they formulate strategy-based rationales for selection that refrain from specifications of, and thus from restrictions to, any particular personality domains (Uher 2008b). For example, the lexical hypothesis proposes the selection of human personality descriptors from the human lexica without confining this selection to any particular domains of behavior and thus of personality differences.

Top-down/etic approaches, in contrast, fail to enable comprehensive selections because they formulate content-based rationales that confine selections to those trait constructs and measures that are imported from other species or populations; thus, they determine a priori the behavioral domains to be studied for personality differences. For this reason, top-down/etic approaches can only reveal evidence for the applicability of trait constructs or measures to other populations within the range of imported personality domains, but not beyond. Yet they may fail to identify population-specific domains of interindividual behavioral variation that those other populations, from which the constructs and measures are imported, do not exhibit. This may result in incomprehensive taxonomic models in which important personality factors are biased or even missing completely (Church 2001; Uher 2008a, b).

For example, top-down/etic approaches based on trait descriptors of the human Five Factor Model yielded different patterning effects both in orangutans (Weiss et al. 2006), and in chimpanzees (King and Figueredo 1997; King et al. 2005; King and Weiss 2011). However, these species differences could be established only within the scope of personality domains described by these trait descriptors, but not beyond. For example, in adaptation to their ecological niches, orangutans and chimpanzees may have developed species-specific domains of interindividual behavioral variation that humans may not show, and that can therefore not be identified with top-down approaches from human personality descriptors. Yet, in chimpanzees, this top-down approach was more comprehensive than a top-down approach based on trait measures originally selected for rhesus macaques (Murray 1998), which could yield only half of the personality domains shown with a top-down/etic approach based on descriptors of human personality factors. This illustrates the substantial impact selection procedures have on the comprehensiveness of empirically identified trait taxonomies.

Bottom-up/emic approaches, in contrast, study personality "as from inside the system" (Pike 1967, p 120). Thus, if they are applied systematically, they enable comprehensive selections of behavioral domains that can be studied for personality differences. Moreover, because they rely on population-specific trait constructs and measures, bottom-up/emic approaches ensure that the behaviors studied for

interindividual differences reflect behaviors that actually occur in natural settings; that is, they ensure that the thus derived personality constructs are ecologically valid. Replications of similar personality factors across different populations on the basis of population-specific trait measures provide strong evidence for the universality of the behavioral variation they explain (Church 2001). For example, the lexical bottom-up approach was carried out in human populations speaking English (Allport and Odbert 1936; Goldberg 1990), German (Angleitner et al. 1990), and Dutch (Hofstee et al. 1981), amongst others. Five strongly similar factors emerged in English and German, whereas two additional factors were shown in the Dutch.

Ecological validity cannot be ensured by top-down/etic approaches, however, because they import specific trait constructs and measures from other populations, and thus study personality "as from outside of a particular system" (Pike 1967, p 120). They may sometimes force constructs and measures on a species that may not be applicable to that species (Gosling et al. 2003, p. 283), and that consequently lack ecological validity (Uher 2008a, b; Uher and Asendorpf 2008).

The potentials for comprehensive selections of ecologically valid personality constructs are also limited in nomination approaches (e.g., Stevenson-Hinde and Zunz 1978) that are likewise based on content-based selection strategies. Being nonconspecific outsiders, we have only limited access to nonhuman species; intuitive nominations by a few knowledgeable informants therefore run the risk of overlooking important individual differences that are not salient to human observers or that do not match their implicit personality theories.

Eclectic approaches try to increase comprehensiveness of selections by combining findings and methodologies from different approaches. They mostly rely on top-down/etic approaches from trait constructs developed for different species (e.g., Rouff et al. 2005) that are sometimes also complemented with expert nominations (Freeman et al. 2011). The comprehensiveness of this content-based selection strategy depends not only on the existing knowledge about other species and the trait constructs that have been developed for them, but in particular on the rationales used to select trait constructs across species and studies, and to merge diverse constructs in order to eliminate redundancies. Yet these rationales are rarely described explicitly (Uher 2008a, b).

Adaptive approaches, by contrast, may be suited for comprehensive selections of ecologically valid traits constructs; but to my knowledge, they have not yet been applied to nonhuman primates (Uher 2008a).

3.3.2 The Behavioral Repertoire × Environmental Situations Approach

In human personality research, systematic bottom-up/emic approaches from the lexica proved to be extremely useful to establish ecologically valid and comprehensive trait taxonomies. The behavioral repertoire × environmental situations approach (Uher 2008a, b) can be considered an alternative systematic bottom-up approach

that derives trait constructs from inside the behavioral and ecological system of a population. Its rationale is grounded in trait psychology and conceives personality differences as interindividual differences in intraindividually stable patterns of conditional probabilities to display particular categories of behaviors in particular categories of environmental situations. Consequently, the approach proposes compiling all important behavioral categories from the known behavioral repertoires of populations (usually species), and plotting them systematically against all situational categories in which they are typically displayed. The resulting behavior-situation units are used to derive hypothetical personality constructs that are then studied empirically for temporally consistent interindividual variability. These empirical analyses are essential since the trait constructs are construed only theoretically; they need not reflect empirical domains of interindividual variability in the studied population. If individuals show no variability or temporal consistency therein, the particular construct is discarded. Finally, these trait constructs are analyzed for intercorrelations and reduced to a few factors in order to derive a structural personality model that describes the studied population (Uher 2008a, b).

Similarly, Gosling et al. (2003, p 283) postulated that "to ensure comprehensiveness, the range of personality traits studied in a species must fully represent the behavioral repertoire of that species." The behavioral repertoire × environmental situations approach fulfills this requirement and extends beyond it. First, the behavioral repertoire approach considers not only the behavioral repertoire, but also the categories of environmental situations in which certain behaviors are typically displayed. This crucial element is inherent to the rationale of the approach. It is derived from trait psychological findings of consistent interactional patterns between individual and situational features. Explicit incorporation of the individuals' environments also opens up connections to ecological and evolutionary perspectives on personality differences (Uher 2008b).

Second, the behavioral repertoire × environmental situations approach generates theoretical constructs and not trait measures. It is not the behavioral categories compiled in the review that are studied, but theoretical constructs derived from a broad range of behavioral and situational categories. For this reason, the approach can consider behavioral categories of various types and functions that would not fit into the homogeneous and disjunctive categories of one single ethogram. This is, however, necessary to actualize a comprehensive approach. Once trait constructs are generated, measures for empirical investigation are systematically selected, which also helps to keep their number manageable for empirical studies (Uher 2008b).

Third, instead of studying behavior from scratch, the approach generates trait constructs from behavioral and situational categories of known meaning and function. It capitalizes on the expertise behavioral sciences have gained on the behavior of the average individual of the study population, and searches systematically for consistent variation among individuals therein (Uher 2008a, b).

The behavioral repertoire × environmental situations approach has already been applied to the great apes species (Uher 2008a, b; Uher and Asendorpf 2008; Uher et al. 2008). The behavioral and situational categories that were cataloged on a broad and general level for each of these closely related species were strikingly similar.

They were therefore pooled to generate trait constructs that are likely applicable to all great ape species. For initial empirical tests in a small sample of captive individuals, behavioral and situational categories that can only be observed in the wild were excluded. Furthermore, traits involving the same behavioral categories, but more specific situational categories were subsumed within one broader trait construct. For example, arousability in social vs. nonsocial situations was subsumed within one arousability construct. This trait generation procedure yielded 19 qualitatively distinct potential trait constructs (listed in Fig. 3.1). Methodological studies in a sample of 20 zoo-housed great apes, among them the Uher et al. (2008) study already discussed earlier, provide initial empirical evidence for stable interindividual differences that are described by these trait constructs (Uher 2008a, b; Uher and Asendorpf 2008).

The behavioral repertoire × environmental situations approach yielded substantial empirical evidence for temporally reliable interindividual differences in very similar trait domains as those shown by top-down/etic approaches in these species – and could also show some further trait domains beyond. These include food orientation, friendliness to youngsters, or sexual activity that are important for great apes, but that have been excluded during the development of the human Big Five factors (see e.g., Schmitt and Buss 2000). These findings emphasize that top-down/etic approaches may permit first explorations of so far unstudied species, but they ultimately require empirical convergence to bottom-up/emic findings to validate the comprehensiveness and ecological validity of their trait constructs for each particular species; for detailed discussions see Uher (2008a, b).

3.4 Methods of Measurement for Primate Personality

Personality constructs can be measured with various methods. The choice of assessment method is thereby independent of the methodological approach; these are two separate meta-theoretical steps (Uher 2008b). This means that trait constructs of human personality derived with lexical bottom-up approaches can be measured not only with lexical trait descriptors, but also with ethological measures of behavior. And vice versa, constructs derived with the behavioral repertoire x environmental situations approach can also be measured with ratings on lexical trait descriptors as I will show now.

3.4.1 The Diversity of Assessment Methods

In nonhuman research, methods of personality assessment are often classified into two groups with coding or ethological behavior observations labeled as objective methods on the one hand, and ratings labeled as subjective methods on the other hand (Gosling 2001, 2008; Capitanio 2004; Freeman et al. 2011). But in fact, methods

of personality assessment span a continuum from records of single behavioral acts to ratings of adjectives as abstract personality descriptors, with methods utilizing elements of both, such as act frequency ratings (Borkenau et al. 2004) or behavior-descriptive verb ratings (Uher and Asendorpf 2008), in the middle.

Ethological methods of behavior measurement (see Altmann 1974; Lehner 1996) are close to the behavioral act pole of this continuum. Since they are based on direct observations of behavior, they seem to suggest greater objectivity than ratings. But no observation of behavior is without abstraction. The observer has to group behavioral events into classes by abstracting properties that recur in more than one event across different levels of behavioral complexity ranging from single muscle movements to more abstract behavior categories (Lehner 1996), and this is inevitably a subjective process. Thus, although the well-defined, homogeneous and independent categories of ethograms should minimize the scope left for subjective decisions, behavioral observations always do have subjective components.

Whereas subjectivity may be lowest in ethological observations, it is highest in abstract ratings that are close to the opposite pole of the objective-subjective continuum of assessment methods. Ratings rely on human ability to differentiate individuals reliably, to perceive individual behavior, to recall observations from multiple occasions in different situations over time, to aggregate this information mentally, and to express overall judgment on predefined sets of personality descriptors (so-called items) in standardized psychometric scales (Funder 1999; Uher and Asendorpf 2008). Hence, all methods of personality measurement are eventually based on observable behavior. They differ only in the *degree* of subjectivity with which they make it possible to capture interindividual behavioral variation.

Not just any measure is per se a useful measure of personality constructs. Personality measures must differentiate well and reliably among individuals, that is, they must have high discriminatory power (Kline 2000; see also Fairbanks and Jorgensen 2011). This must also be shown for rating data; independent raters must agree substantially and provide reliable distinctions between individuals. Interrater reliability can be determined for both the rank order of the individuals on a given personality descriptor (variable-oriented view), and the individual profiles across multiple items (individual-oriented view).

Ratings are often assumed to imply stability in the targets' behavior since they are derived from mental aggregations by the judges. But human observers tend to overestimate stability (Uher and Asendorpf 2008). For instance, a few observations of extreme instances of behavior, such as strong aggression, can bias observers to assume overall high aggressiveness. When later observing mild aggression by the same animal, observers may judge this as an instance of high aggression. These biases can occur even in repeated observations of concrete behaviors, but are more marked in global judgments based only on intuitive aggregation of observed earlier behavior. Such biases may become particularly problematic when observation time is limited. Establishing test–retest reliability for rating measures is thus as important as it is for behavior measures. I now illustrate analyses of interrater and test–retest reliability with rating data obtained with the Great Ape Personality Inventory (GAPI).

3.4.2 Ratings on Behavior-Descriptive Verbs and Trait-Adjectives: The Great Ape Personality Inventory (GAPI)

The GAPI is a psychometric instrument to assess in captive Great Apes (bonobos, chimpanzees, gorillas, and orangutans) personality traits that were derived with the behavioral repertoire × environmental situations approach (see above, Uher 2008a, b). It is available in two complementary formats that are useful for validation. The behavior-descriptive verb form (*GAPI-B*) describes observable, trait-indicating behaviors in circumscribed situations using verbs only. Food orientation, for example, is described with "When there is food, *Name* is (often) quickly on the spot." Thirty-four items were constructed, of which ten are reversed in their meaning to reduce the effects of response sets. The trait adjective form (*GAPI-A*) describes the trait constructs with single trait adjectives in everyday language such as "*Name* is (very) gluttonous." None of the 17 items is reversed in meaning. English translations of the original German items are provided in Tables 3.1 and 3.2.

Table 3.1 Great ape personality inventory – behavior-descriptive verb items (GAPI-B)[a]

Personality trait construct	Items *GAPI* – behavior-descriptive verb items	Code	Interrater reliability		Temporal reliability
			$ICCt_1$	$ICCt_2$	*r*
Aggressiveness to humans	*Name* (often) jumps at the grate or window when persons stay in front of it	AG2	0.86	0.78	0.94***
	Name (often) spits or throws objects from the enclosure	AG3	0.84	0.88	0.94***
	Name (often) tries to scratch persons through the grate	AG4	0.92	0.69	0.82***
Anxiousness	When *name* is alone in a room he/she (often) moves about continuously, and sometimes has diarrhea	AX2	0.79	0.90	0.95***
	When one comes close to the grate near to *name*, he/she (often) shies away quickly	AX3	0.06	0.25	0.86***
Arousability	Prior to the feeding, *name* (often) moves about a lot	AR2	−0.21	0.62	0.59**
	When being fed, *name* (often) makes many sounds	AR3	−0.17	0.82	0.92***
Curiosity	*Name* (often) touches new objects, such as enrichment items, at great length	CU2	0.76	0.60	0.91***
	Confronted with novel food, *name* (mostly) ignores it	CU3	0.88	0.78	0.61**
Distractibility	When *name* is busy with something, he/she (often) disrupts his/her activity as soon as something else is going on	DI2	0.78	0.73	0.72***

(continued)

Table 3.1 (continued)

Personality trait construct	Items *GAPI* – behavior-descriptive verb items	Code	Interrater reliability		Temporal reliability
			$ICCt_1$	$ICCt_2$	*r*
Dominance	In the group, *name* is (most often) the first to get to the food	DO2	0.96	0.85	0.98***
	Name is (most often) the last to get to the food	DO3	0.98	0.82	0.96***
Food orientation	When there is food, *name* is (often) quickly on the spot	FM2	0.79	–0.50	0.69***
	Between feeding times, one (hardly ever) sees *name* eating	FM3	0.69	0.70	0.64**
Friendliness to humans	When called, *name* (often) comes to the grate closely	FR2	0.89	0.27	0.20
	(At times), *name* even allows close contact with humans	FR3	0.84	0.43	0.88***
Friendliness to conspecifics	*Name* (often) grooms other group members	FR5	0.69	0.82	0.90***
	Name has (hardly ever) body contact with other group members	FR6	0.71	0.86	0.89***
Friendliness to youngsters	*Name* spends (a lot of) time with youngsters	CH2	0.75	0.96	0.91***
	Name (often) plays with youngsters	CH3	0.57	0.49	0.89***
Gregariousness	*Name* (often) withdraws from his/her conspecifics in the indoor or outdoor enclosure	GR2	0.94	0.90	0.89***
	Name sits together with his/her conspecifics (a lot)	GR3	0.93	0.89	0.88***
Impulsiveness	When he/she does not get his /her food immediately, *name* (often) quickly knocks at the grate or window	IM2	0.04	0.42	0.68***
	Name (often) waits calmly until it is his/her turn to get his/her food	IM3	–0.28	0.68	0.89***
Persistency	With dealing with enrichment materials, *name* (often) gives up easily	PE2	0.11	0.64	0.79***
	Name can keep him-/herself busy with something (for a long time)	PE3	0.41	0.56	0.94***
Physical activity	In the indoor or outdoor enclosure, *name* keeps walking or brachiating (most of the time)	AC2	0.93	0.93	0.98***
	(Most of the time), *name* is sitting or lying	AC3	0.96	0.90	0.90***

(continued)

Table 3.1 (continued)

Personality trait construct	Items *GAPI* – behavior-descriptive verb items	Code	Interrater reliability		Temporal reliability
			$ICCt_1$	$ICCt_2$	*r*
Playfulness	*Name* (often) plays on his/her own with objects such as enrichment items	PL2	0.76	0.42	0.88***
	Name (rarely) plays with adolescent or adult members of the group	PL3	0.82	0.77	0.79***
Sexual activity	*Name* (often) establishes sexual contact with his/her conspecifics	SX2	0.77	0.96	0.96***
	Name (often) stimulates him-/herself sexually	SX3	0.79	0.72	0.90***
Vigilance	*Name* (often) notices small changes in the cages or enclosures quickly	VI2	−0.69	0.17	0.85***
	Name (often) watches everything around him/her very closely	VI3	−0.12	0.49	0.83***
Mean			0.72	0.74	0.88

Note: these are translations of the original German items with which the presented data were collected. The German items can be obtained from the author. Some items can be reversed in meaning depending on whether they are used as *agreement* scales from (1) *strongly disagree* to (5) *strongly agree*, for which the statements of frequency given in parenthesis should be included in the item text, or as *frequency* scales from (1) *hardly ever* to (5) *very often* on items presented without the frequency quantifying expressions provided in parentheses. Variable-oriented interrater reliability rated with the *great ape personality inventory* (*GAPI*) – *behavior-descriptive verb items* (B) in test periods t_1 and t_2 was computed with *ICC* (3,*k*). It depicts reliability of the mean ratings on the basis of $k = 4–5$ independent raters per ape (Shrout and Fleiss 1979). For analyses of test–retest reliability, the scores were aggregated over all raters within each rating period. Variable-oriented test–retest reliability of these aggregated scores over the 5 weeks between rating periods t_1 and t_2 was computed with Pearson correlation *r*. $^{***}p < 0.001$, $^{**}p < 0.01$. Mean reliability scores across the 34 items were computed with *r*-to-*Z* transformation

[a]For captive samples

Ten keepers rated the same 20 individuals studied behaviorally by Uher et al. (2008) on a computer-based interface. On each format, they specified their level of agreement with the statements given in the items on five-point Likert agreement scales from (1) strongly disagree to (5) strongly agree. For this reason, the items contained statements of frequency (such as "often" or "hardly" in *GAPI-B*) or of degree of intensity (such as "very" in *GAPI-A*; given in parentheses in Tables 3.1 and 3.2). Alternatively, ratings could be indicated on frequency scales from (1) hardly ever to (5) very often on items presented without the frequency quantifying expressions provided in parentheses. This could facilitate understanding of the items, but would hinder inversions of item meanings, thus increasing probabilities of response sets.

For comparisons among methods, ratings were scheduled to parallel the behavioral data collection of the Uher et al. (2008) study. All individuals were rated twice by four to five raters, with an interval of 5 weeks (for details see Uher and Asendorpf 2008). This design allowed analyses of interrater reliability for each data collection

Table 3.2 Great ape personality inventory – trait adjective items (GAPI-A)[a]

Personality trait construct	Items *GAPI* – trait adjective items	Code	Interrater reliability		Temporal reliability
			$ICCt_1$	$ICCt_2$	*r*
Aggressiveness to humans	To humans, *name* is (very) aggressive	AG1	0.90	0.61	0.80***
Anxiousness	*Name* is (very) anxious	AX1	0.73	0.73	0.81***
Arousability	*Name* is (quickly) excited	AR1	0.82	0.80	0.89***
Curiosity	*Name* is (very) curious	CU1	0.47	0.74	0.87***
Distractibility	*Name* is (very) distractible	DI1	0.75	0.65	0.80***
Dominance	*Name* is (very) dominant	DO1	0.97	0.94	0.98***
Food orientation	*Name* is (very) gluttonous	FM1	0.86	0.83	0.83***
Friendliness to humans	To humans, *name* is (very) friendly	FR1	0.54	0.82	0.89***
Friendliness to conspecifics	To her conspecifics, *name* is (very) friendly	FR4	0.78	0.61	0.93***
Friendliness to youngsters	To youngsters, *name* is (very) friendly	CH1	0.03	0.35	0.67***
Gregariousness	*Name* is (very) gregarious	GR1	0.89	0.88	0.91***
Impulsiveness	*Name* is (very) impulsive	IM1	0.82	0.50	0.81***
Persistency	*Name* is (very) persistent (such as with enrichment materials)	PE1	0.39	0.73	0.90***
Physical activity	*Name* is physically (very) active	AC1	0.98	0.92	0.92***
Playfulness	*Name* is (very) playful	PL1	0.95	0.92	0.91***
Sexual activity	*Name* is sexually (very) active	SX1	0.78	0.95	0.96***
Vigilance	*Name* is (very) vigilant	VI1	–0.28	0.53	0.73***
Mean			0.79	0.79	0.88

Note: these are translations of the original German items with which the presented data were collected. The German items can be obtained from the author. Variable-oriented interrater reliability rated with the *great ape personality inventory* (*GAPI*) – *trait adjective items* (A) in test periods t_1 and t_2 was computed with *ICC* (3,*k*). It depicts reliability of the mean ratings on the basis of $k = 4–5$ independent raters per ape (Shrout and Fleiss 1979). For analyses of test–retest reliability, the scores were aggregated over all raters within each rating period. Variable-oriented test–retest reliability of these aggregated scores over the 5 weeks between rating periods t_1 and t_2 was computed with Pearson correlation *r*. $^{***}p < 0.001$, $^{**}p < 0.01$. Mean reliability scores across the 17 items were computed with *r*-to-*Z* transformation

[a]For captive samples

period, and analyses of test–retest reliability between periods. Interrater reliability was substantial in both variable-oriented and individual-oriented analyses. In the first rating period, the mean variable-oriented reliability among the $k = 4–5$ independent raters per ape as indicated by ICC(3,*k*) (Shrout and Fleiss 1979) was 0.72 for behavior-descriptive verbs and 0.79 for trait adjectives. Mean individual-oriented interrater agreement was ICC(3,*k*) = 0.84 for behavior-descriptive verbs, and 0.85 for trait adjectives. Results on the item level are given in Tables 3.1 and 3.2 those on the individual level are given in Table 3.3, separately for the two periods of data collection.

Table 3.3 The subjects and individual-oriented analyses of interrater reliability, test–retest reliability, and validity of personality profiles rated with the GAPI

Subjects				Interrater reliability *GAPI*				Temporal reliability *GAPI*		Validation		
				Behavior-descriptive verb items *B*		Trait-adjective items *A*		Behavior-descriptive verb items *B*	Trait-adjective items *A*	Cross-method coherence between		
Species	Name	Age	Sex							B–A	E–B	E–A
				$ICCt_1$	$ICCt_2$	$ICCt_1$	$ICCt_2$	*r*	*r*	*r*	*r*	*r*
Bonobo	Joey	22	M	0.76	0.86	0.85	0.87	0.87***	0.85***	0.53*	0.60**	0.22
	Kuno	8	M	0.92	0.91	0.92	0.94	0.97***	0.97***	0.34	0.94**	0.31
	Limbuko	9	M	0.76	0.85	0.85	0.84	0.93***	0.83***	0.55*	0.40	−0.01
	Ulindi	11	F	0.87	0.80	0.86	0.80	0.93***	0.92***	0.66#	0.49*	0.39
	Yasa	7	F	0.68	0.77	0.88	0.89	0.83***	0.97***	0.51*	0.70**	0.28
Chimpanzee	Dorien	24	F	0.65	0.80	0.19	0.70	0.83***	0.71**	0.47	0.69**	0.35
	Fraukje	28	F	0.76	0.82	0.86	0.80	0.91***	0.90***	0.87***	0.51*	0.34
	Frodo	11	M	0.76	0.85	0.80	0.76	0.79***	0.90***	0.28	0.61**	−0.05
	Robert	29	M	0.85	0.82	0.76	0.84	0.89***	0.90***	0.44#	0.62**	0.06
	Sandra	11	F	0.70	0.82	0.82	0.84	0.88***	0.96***	0.70**	0.79***	0.54*

Gorilla	Bebe	25	F	0.80	0.76	0.88	0.90	0.89***	0.92***	0.62**	0.51*	0.27
	Gorgo	23	M	0.77	0.81	0.50	0.66	0.96***	0.93***	0.58#	0.74**	0.22
	N'diki	27	F	0.75	0.78	0.83	0.75	0.87***	0.93***	0.62**	0.28	0.46#
	Ruby	7	F	0.86	0.78	0.89	0.93	0.90***	0.94***	0.31	0.73**	−0.06
	Viringika	9	F	0.77	0.78	0.82	0.87	0.84***	0.91***	0.73***	0.68**	0.40
Orangutan	Bimbo	24	M	0.84	0.84	0.87	0.85	0.95***	0.91***	0.63#	0.54*	0.05
	Dokana	16	F	0.91	0.93	0.93	0.91	0.96***	0.96***	0.72**	0.80**	0.77**
	Dunja	31	F	0.85	0.87	0.87	0.89	0.95***	0.96***	0.34	0.58*	0.13
	Padana	7	F	0.86	0.87	0.84	0.88	0.93***	0.89***	0.47#	0.46#	0.48#
	Pini	16	F	0.90	0.89	0.83	0.85	0.95***	0.92***	0.55*	0.79***	0.47#
Mean				0.81	0.84	0.83	0.85	0.91	0.92	0.57	0.67	0.30

Note: details on the subjects' species, their age in years, and sex (F = female, M = male). Reliability of the individuals' personality profiles rated with the *great ape personality inventory* (*GAPI*) *behavior-descriptive verb items* (B) and *trait adjective items* (A) was computed with *ICC* (3,*k*) separately for the two non-overlapping test periods t_1 and t_2. It depicts reliability of the mean ratings on the basis of $k = 4–5$ independent raters per ape (Shrout and Fleiss 1979). For analyses of test–retest reliability, the scores were aggregated over all raters within each rating period. Individual-oriented test–retest reliability of these aggregated scores over the 5 weeks between rating periods t_1 and t_2 was computed with Pearson correlation r. $^{***}p < 0.001$, $^{**}p < 0.01$. Coherence between the individuals' personality profiles across 17 traits rated with the *GAPI – behavior descriptive verb items* (B), with the *GAPI – trait adjective items* (A), and with ethological behavior measures (E) obtained from observations in 14 laboratory tests and group situations was computed with Pearson correlations r. $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\#}p < 0.10$. To further increase the reliability of the personality profiles obtained with each method, they are based on data that were each aggregated on the trait level across the two studied time periods spanning about 6 weeks. Corresponding data from variable-oriented analyses are reported in Uher et al. (2008), and Uher and Asendorpf (2008)

Since ratings showed high interrater reliability, mean rating scores were calculated across keepers within each rating period. Test–retest reliabilities between these averaged ratings were substantial. Over 5 weeks, variable-oriented correlations were $r = 0.88$ for both behavior-descriptive verb and trait adjective items; individual-oriented correlations were $r = 0.91$ for behavior-descriptive verb and 0.92 for trait adjective items. Results on the item level are given in Tables 3.1 and 3.2, and those on the individual level are given in Table 3.3. Comparisons of test–retest reliability scores between different personality measures showed that those obtained with ratings were significantly higher than those obtained with ethological methods; the effect sizes were large ranging from $d = 0.73$ to 0.91 (Uher and Asendorpf 2008). These results should be kept in mind when interpreting temporal reliability or temporal stability of personality differences based on rating methods.

3.4.3 Validation in Personality Research

Personality measures must not only be reliable, they must also be valid. That is, it must be shown that they measure what they are supposed to measure. Establishing empirical validity is crucial for research on theoretical constructs such as personality traits. The central concern is thus to link a theoretical concept with empirical findings. This is the purpose of validation through nomological networks. A nomological network includes a theoretical framework that represents the basic features of the trait construct in question, an empirical framework how this shall be measured, and specification of the interrelationships among and between these two frameworks (Cronbach and Meehl 1955). For example, if one is interested in curiosity, a weak approach is to study it with only one method, whether by rating or by ethological measure. A stronger approach is to do both and to show coherence between the different measures of the construct of curiosity. Converging evidence from different methods establishes a strong case of construct validation for the studied personality construct (Cronbach 1988).

I illustrate the use of nomological networks with data from great apes. I analyzed the construct validity of personality traits derived with the behavioral repertoire × environmental situations approach in these species (Uher 2008a, b) with three different assessment methods. That is, for each trait construct, I specified a priori several ethological behavior measures (Uher et al. 2008), two behavior-descriptive verb items, and one trait adjective item that theoretically should reflect that construct well. These measures span a nomological network around each trait construct. For most traits, the theoretical relations among these measures could be substantiated empirically. The mean variable-oriented correlation across 17 trait constructs between behavior-descriptive verb ratings and trait adjective ratings was $r = 0.71$; between behavior-descriptive verb ratings and composite ethological behavior measures it was $r = 0.56$; and between trait adjective ratings and composite ethological behavior measures it was $r = 0.35$ (Uher and Asendorpf 2008). Mean individual-oriented correlations across 20 individual personality profiles were

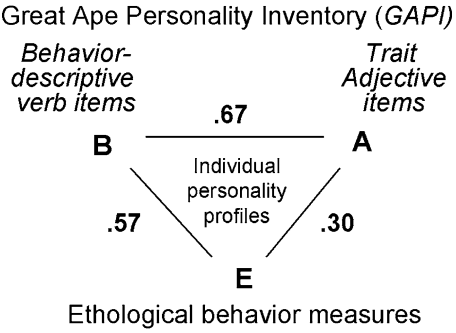

Fig. 3.3 Individual-oriented coherence across 20 individuals (mean Pearson correlations *r* computed with *r*-to-*Z* transformation) among individual personality profiles rated with the great ape personality inventory (*GAPI*) – *behavior-descriptive verbs* (B) and the *GAPI – trait adjectives* (A), and measured with ethological methods of behavior measurement (E) in a series of 14 laboratory tests and group observations. To further increase the reliability of the personality profiles obtained with each method, they are based on data that were each aggregated on the trait level across the two studied time periods spanning about 6 weeks. The ethologically measured behavior profiles of four individuals were incomplete since the subjects could not be tested in the laboratory

virtually identical (Fig. 3.3). This established substantial evidence for the construct validity of personality trait constructs derived with the behavioral repertoire × environmental situations approach.

These studies are also useful to explain the processes of validating psychometric instruments for personality ratings in nonhuman species. Standard inventories of human personality are based on (1) a theoretical foundation. They are developed using iterative procedures of empirical testing and statistical item selections. For the resulting instruments, empirical evidence for sufficient (2) interrater agreement and (3) test–retest reliability for each single item, (4) validity for each single personality construct as well as their (5) empirical intercorrelations and factor structure are routinely shown in large samples, but these characteristics are generally taken for granted in later applications (Kline 2000). These standard criteria are documented in application manuals together with (6) norm distributions for specific reference populations.

Surprisingly, these essential and well-established methodological foundations of instrument development have received only very little attention in primate personality research. The *GAPI* is one of the first published primate personality inventories for which the first four of these six essential standard steps of instrument development have been accomplished. Except for top-down/etic approaches from rating items of the human Five-Factor Model (King and Figueredo 1997; Weiss et al. 2006), which are grounded in phylogenetic theory, to my knowledge, no other rating list published to date for taxonomic personality research is based on a theoretical foundation. Interrater reliability is almost always analyzed, but test–retest reliability is rarely studied (for exceptions see McGuire et al. 1994; Stevenson-Hinde et al. 1980; Uher and Asendorpf 2008).

First steps towards validation have already been made for some rating lists by showing empirical relations to single behavior measures (McGuire et al. 1994;

Capitanio 1999; Pederson et al. 2005; Kuhar et al. 2006). However, the behavior measures were often selected without a priori specification of their theoretical relationships to the studied trait constructs. Many of these behaviors were selected from ethograms that are used for research questions other than personality differences. As already noted earlier, however, every behavior is not per se a useful measure of personality. Since most of these studies failed to analyze the test–restest reliability of their behavioral measures, it remains unclear whether they are sufficiently aggregated to represent in fact reliable personality measures. Unless test–restest reliability is shown for behavioral measures, coherence with rating measures, and thus validation, may be compromised. This could explain why many studies show only low to moderate correlations between ratings and ethological behavior measures of nonhuman primate personality.

The use of nomological networks in the Uher and Asendorpf (2008) study, and empirical test–restest reliability of all obtained measures, that is of ethological personality measures and two kinds of personality rating measures, allow systematic analyses of the validity of the GAPI. Item analyses are important since it is the items that activate the raters' pertinent knowledge, that initiate their mental assessment processes, and that provide the frameworks in which the raters can indicate their resulting judgments (Funder 1999; Uher and Asendorpf 2008). Item analyses are particularly relevant for inherently anthropocentric trait adjective items. Their use for personality ratings in humans is theoretically (Goldberg 1990) and empirically well founded (Kenrick and Funder 1988), but evidence for their validity in nonhuman species is rarely provided, despite their popularity. Yet without systematic validation, the behaviors they actually refer to in particular species remain unclear as well as what they are actually measuring (Uher 2008a, b; Uher and Asendorpf 2008).

Trait adjectives can have implicit connotations for raters that are not obvious from their general meaning. For example, in great apes, "friendly to his/her conspecifics" was surprisingly uncorrelated with both behavior-descriptive verb ratings and ethological behavior measures of grooming and body contact. This finding could indicate that keepers base their judgments of individuals as "friendly" not on prosocial behaviors, such as grooming, but instead on low aggression. This would have significant implications for predictions of behavior in particular situations, such as in group introductions, because low aggressiveness may not necessarily imply high prosociality. Differences in interpretation like these, which are neither obvious nor intended, are obscured by items that complement trait adjectives with "clarifying behavioral definitions" as frequently used for primate ratings, such as defining "gentle" with "responds to others in an easy, kind manner" (Stevenson-Hinde and Zunz 1978; McGuire et al. 1994; King and Figueredo 1997; Weiss et al. 2006). Separate analyses of trait adjectives, and their supposed behavioral definitions, are thus important for validation (Uher and Asendorpf 2008).

Trait adjectives require large inferences from observable behavior, and may therefore be prone to anthropomorphic interpretations of behavior. Behavior-descriptive verb items, in contrast, are less inferential and less susceptible to biases and subjectivity than trait adjective items since they require the raters to focus on specific, perceivable behaviors. Validation analyses of the GAPI show that coherence

with ethological behavioral measures of personality is substantially higher for behavior-descriptive verbs than for trait adjectives (see Fig. 3.3; Tables 3.1–3.3). This may be because behavior-descriptive verbs are both behaviorally based, whereas trait adjectives as abstract personality descriptors may have broader predictive ranges of behaviors and situations. However, a study on human personality could not clearly support the hypothesis that trait adjectives generally refer to more exemplars than verbs (Borkenau and Müller 1991). Rather, the relation between the grammatical form of personality-descriptive categories and the number of their exemplars was mediated by category breadth. If category breadth was held constant, grammatical form correlated significantly with the rated trait prototypicality of their exemplars. That is, verbs describe more accurately how individuals are actually behaving than adjectives (Borkenau and Müller 1991).

The empirical results presented in this chapter and in the Uher and Asendorpf (2008) study square nicely with these findings. They underscore the particular utility of non-trait-adjective rating methods, such as behavior-descriptive verb or act-frequency ratings, which combine the greater accuracy of behavior prediction with the economy of rating methods.

It is obvious that ratings constitute economic methods of personality assessment (Vazire et al. 2007; Gosling 2008; Freeman et al. 2011), but they do so only if their validity is evidenced empirically. As I have shown, psychometric validation requires substantial empirical and statistical work that in nonhuman species ultimately includes coherence with observable behavior. Thus initially, ratings are much more labor-consuming methods of personality assessment than ethological methods. But as validated psychometric instruments, they allow economic measurements of personality. For the GAPI, four of six essential steps of standard instrument development have already been accomplished. Further steps require empirical studies in larger samples to analyze the species' factor structures and their norm distributions. They could also include psychometric analyses of larger item pools for iterative processes of statistical item selections.

In conclusion, there is no single method of personality assessment that is generally inferior or superior to others. The question of method selection should therefore not be polarized by premature recommendations (as in Vazire et al. 2007; Gosling 2008; Freeman et al. 2011) that obscure the diversity of assessment methods and the important functions this very diversity serves for construct validation. Instead, the advantages and disadvantages of the different methods of measurement should be weighed selectively for their relevance to the particular research questions at hand (Uher 2008a, b).

3.4.4 Establishing Comparability of Trait Constructs Across Populations

Populations, such as species, can also show population-specific behaviors that are not shown in other populations. This must be considered when personality variation is compared among populations. Cross-population comparisons presuppose

comparability of trait constructs even if they are measured with different behaviors. A first step is analysis of functional equivalence of behaviors used to measure a trait construct (Mehta and Gosling 2008). Comparability analyses of meaning and functions of behaviors have been established in ethology and rely on fine-grained contextual prepost analyses of behavioral sequences (Preuschoft 1992; Preuschoft and van Hooff 1995).

Yet functional equivalence of behavioral measures alone is insufficient to conclude that personality constructs are comparable across populations. Personality variation can be compared only on the construct level, not on the level of single measures (Uher 2008b); comparability of trait constructs therefore has to be established empirically as structural equivalence. Methodologies for statistical comparison of factorial structures of functionally equivalent, yet nonidentical trait measures across different population levels have been established in cross-cultural research (Vijver and Poortinga 2002). They can be generalized to other population comparisons such as among species (for details see Uher 2008b).

Since all ratings necessarily rely on human language, researchers using trait adjective ratings are tempted to assume that identical items also imply comparability across the different species to which they are applied (Weiss and Adams 2008). But because trait adjectives can have fairly different implicit connotations in other species, their "functional" equivalence has to be established first through empirical convergence with behavioral measures. That factorial structures of personality constructs obtained with identical items can differ among species has already been shown descriptively (King and Figueredo 1997; Weiss et al. 2006). But so far, structural equivalence of such factors has been analyzed statistically only between two different populations of captive chimpanzees (Weiss et al. 2007); statistical analyses of their structural equivalence or nonequivalence across species are still pending.

3.5 Conclusions

Human personality psychology provides a rich and solid foundation of theoretical concepts, methodological approaches, and methods of assessment with unquestionable suitability for nonhuman primate personality research. Many concepts and methodologies for within-population research are directly applicable to nonhuman primates. Those established for cross-cultural comparisons of human population can be generalized systematically to comparisons of nonhuman populations including species. There is much for us to learn from human personality psychology; its knowledge and experiences in solving many puzzling research issues can give nonhuman personality research a competitive edge to head for new advances in the near future.

Acknowledgments I am grateful to the editors and to Wendy Johnson, Jochen Fahrenberg, and Jens Asendorpf for valuable comments on the manuscript. I thank the zookeepers at the Wolfgang Köhler Primate Research Center in Leipzig, Germany, for their kind cooperation and for rating the apes; Josep Call from the Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany, for supporting this study; and Josefine Kalbitz for recording behavior for the reliability analyses.

References

Allport GW (1937) Personality: A psychological interpretation. Macmillan, New York
Allport GW, Odbert HS (1936) Trait names: A psycholexical study. Psychol Monogr 47:1
Altmann J (1974) Observational study of behavior: Sampling methods. Behaviour, 49:227–267
Angleitner A, Ostendorf F, John OP (1990) Towards a taxonomy of personality descriptors in German: A psycho-lexical study. Eur J Pers 4:89–118
Asendorpf JB (1988) Individual response profiles in the behavioral assessment of personality. Eur J Pers 2:155–167
Asendorpf JB, van Aken MAG (1999) Resilient, overcontrolled, and undercontrolled personality prototypes in childhood: Replicability, predictive power, and the trait-type issue. J Pers Soc Psychol 77:815–832
Block J (1971) Lives through time. Bancroft, Berkeley
Borkenau P, Müller B (1991) Breadth, bandwidth, and fidelity of personality-descriptive categories. Eur J Pers 5:309–322
Borkenau P, Mauer N, Riemann R et al. (2004) Thin slices of behavior as cues of personality and intelligence. J Pers Soc Psychol 86:599–614
Buss DM (1991) Evolutionary personality psychology. Annu Rev Psychol 42:459–491
Cairns RB, Bergman LR, Kagan J (1998) Methods and models for studying the individual. Sage Publications, Thousand Oaks
Capitanio JP (1999) Personality dimensions in adult male rhesus macaques: Prediction of behaviors across time and situation. Am J Primatol 47:299–320
Capitanio JP (2004) Personality factors between and within species. In: Tierry B, Singh M, Kaumanns W (eds) Macaque societies. Cambridge University Press, Cambridge
Caspi A, Roberts BW (1999) Personality continuity and change across the life course. In: Pervin LA, John OP (eds) Handbook of personality: Theory and research (2nd ed). Guilford Press, New York, NY
Cervone D, Pervin LA (2008) Personality: Theory and research (10th edn). Wiley, Hoboken
Church AT (2001) Personality measurement in cross-cultural perspective. J Pers 69:979–1006
Crawford MP (1938) A behavior rating scale for young chimpanzees. J Comp Psychol 26:79–91
Cronbach LJ (1988) Five perspectives on the validity argument. In Wainer H, Brown HI (eds) Test validity. Erlbaum, Hillsdale
Cronbach LJ, Meehl P (1955) Construct validity in psychological tests, Psychol Bull 52:281–302
De Raad B, Barelds DPH (2008) A new taxonomy of Dutch personality traits based on a comprehensive and unrestricted list of descriptors. J Pers Soc Psychol 94:347–364.
Eysenck HJ (1990) Biological dimensions of personality. In: Pervin L (ed). Handbook of personality theory and research. Guilford, New York
Fairbanks LA, Jorgensen MJ (2011) Objective behavioral tests of temperament in nonhuman primates. In: Weiss A, King JE, Murray L (eds) Personality and temperament in nonhuman primates. Springer, New York
Fleeson W (2004) Moving personality beyond the person-situation debate: The challenge and the opportunity of within-person variability. Curr Dir Psychol Sci 13:83–87
Freeman H, Gosling SD, Shapiro SJ (2011) Comparison of methods for assessing personality in nonhuman primates. In: Weiss A, King JE, Murray L (eds) Personality and temperament in nonhuman primates. Springer, New York
Funder DC (1999) Personality judgment: A realistic approach to person perception. Academic, San Diego
Funder DC (2007) The personality puzzle (4th edn). W.W. Norton & Co, New York
Funder DC (2006) Towards a resolution of the personality triad: Persons, situations and behaviors. J Res Pers 40:21–34
Funder DC, Colvin CR (1991) Explorations in behavioral consistency: Properties of persons, situations, and behaviors. J Pers Soc Psychol 60:773–794

Galton F (1869) Hereditary genius: An inquiry into its laws and consequences. Macmillan, London
Goldberg LR (1990) An alternative "description of personality": The Big-Five factor structure. J Pers Soc Psychol 59:1216–1229
Goodall J (1986) The Chimpanzees of Gombe: Patterns of behavior. Harvard University Press, Cambridge
Gosling SD (2001) From mice to men: What can we learn about personality from animal research? Psychol Bull 127:45–86
Gosling SD (2008) Personality in nonhuman animals. Soc Pers Psych Compass 2:985–1001
Gosling SD, Graybeal A (2007) Tree thinking: A new paradigm for integrating comparative data in psychology. J Gen Psychol 134:259–277
Gosling SD, Lilienfeld SO, Marino L (2003) Personality. In: Maestripieri D (ed) Primate psychology, Harvard University Press, Cambridge, Massachusetts
Hartshorne H, May MA (1928) Studies in the nature of character. Vol 1. Studies in deceit. Macmillan, New York
Hebb DO (1949) Temperament in chimpanzees: I. Method of analysis. J Comp Physiol Psychol 42:192–206
Hofstee WKB, Brokken FB, Land H (1981) Constructie van een Standaard-Persoonlijkheids-Eigenschappenlijst (S.P.E.L.). Nederlands Tijdschrift voor de Psychologie 36:443–452
John OP (1990) The "Big Five" factor taxonomy: Dimensions of personality in the natural language and in questionnaires. In Pervin LA (ed) Handbook of personality: Theory and research. Guilford, New York
John OP, Angleitner A, Ostendorf F (1988) The lexical approach to personality: A historical. review of trait taxonomic research. Eur J Pers 2:171–203
Johnson W (2007) Genetic and environmental influences on behavior: Capturing all the interplay. Psychol Rev 114:423–440
Kenrick DT, Funder DC (1988) Profiting from controversy: Lessons from the person-situation debate. Am Psychol 43:23–34
King JE, Figueredo AJ (1997) The Five-Factor Model plus Dominance in chimpanzee personality. J Res Pers 31:257–271
King JE, Weiss A (2011) Personality from the perspective of a primatologist. In: Weiss A, King JE, Murray L (eds) Personality and temperament in nonhuman primates. Springer, New York
King JE, Weiss A, Farmer KH (2005) A chimpanzee (*Pan troglodytes*) analogue of cross-national generalization of personality structure: Zoological parks and an African sanctuary. J Pers 73:389–410
Kline P (2000) The handbook of psychological testing (2nd ed). Routledge, London
Kuhar CW, Stoinski TW, Lukas KE, Maple TL (2006) Gorilla Behavior Index revisited: Age, housing and behavior. Appl Anim Behav Sci 96:315–326
Lehner PN (1996) Handbook of ethological methods (2nd ed). Cambridge University Press, Cambridge
Leung K, Bond MH (1989) On the empirical identification of dimensions for cross-cultural comparisons. J Cross Cult Psychol 20:133–151
Magnusson D (1988) Individual development from an interactional perspective: A longitudinal study. Erlbaum, Hillsdale
Martin JE (2005) The influence of rearing on personality ratings of captive chimpanzees (*Pan troglodytes*). Appl Anim Behav Sci 90:167–181
Marwitz M, Stemmler G (1998) On the status of individual response specificity. Psychophysiology 35:1–15
Matthews G, Deary IJ, Whiteman MC (2003) Personality traits (2nd Ed). Cambridge University Press, Cambridge
McGuire M, Raleigh M, Pollack D (1994) Personality features in vervet monkeys: The effects of sex, age, social status, and group composition. Am J Primatol 33:1–13
Mehta PH, Gosling SD (2008) Bridging human and animal research: A comparative approach to studies of personality and health. Brain Behav Immun 22:651–661

Mervielde I, Asendorpf JB (2000) Variable-centred and person-centred approaches to childhood personality. In: Hampson SE (ed) Advances in personality psychology, Vol. 1. Psychology Press, Hove

Mischel W (1968) Personality and assessment. Wiley, New York

Mischel W (1977) The interaction of person and situation. In: Magnusson D, Endler NS (eds) Personality at the crossroads: Current issues in interactional psychology. Erlbaum, Hillsdale

Mischel W, Peake PK (1982) Beyond déjà vu in the search for cross-situational consistency. Psychol Rev 89:730–755

Mischel W, Shoda Y (1995) A cognitive-affective system theory of personality: Reconceptualizing situations, dispositions, dynamics, and invariance in personality structure. Psychol Rev 102:246–268

Mischel W, Shoda Y, Mendoza-Denton R (2002) Situation-behavior profiles as a locus of consistency in personality. Curr Dir Psychol Sci 11:50–54

Mischel W, Shoda J, Smith RE (2003) Introduction to personality: Toward an integration (7th ed). Wiley, New York

Murray LE (1998) The effects of group structure and rearing strategy on personality in chimpanzees (*Pan troglodytes*) at Chester, London and Twycross Zoos. Int Zoo Yearb 36:97–108

Murray LE (2002) Individual differences in chimpanzee (*Pan troglodytes*) personality and their implications for the evolution of mind. In: Harcourt D, Sherwood B (eds) New perspectives in primate evolution and behaviour. Otley, Westbury Publishing

Pederson AK, King JE, Landau VI (2005) Chimpanzee (*Pan troglodytes*) personality predicts behavior. J Res Pers 39:534–549

Pike KL (1967) Language as behavior and emic and etic standpoints for the description of behavior. In: Borgatta EF (ed) Social psychology: Readings and perspective. Rand McNally, Chicago

Preuschoft S (1992) 'Laughter' and 'smile' in Barbary macaques (*Macaca sylvanus*). Ethology 91:200–236

Preuschoft S, van Hooff JARAM (1995) Homologizing primate facial displays: A critical review of methods. Folia Primatol 65:121–137

Rouff HJ, Sussman RW, Strube MJ (2005) Personality traits in captive lion-tailed macaques (*Macaca silenus*). Am J Primatol 67:177–198

Rushton JP, Brainerd CJ, Pressley M (1983) Behavioral development and construct validity: The principle of aggregation. Psychol Bull 94:18–38

Schmitt DP, Buss DM (2000) Sexual dimensions of person description: Beyond or subsumed by the big five? J Res Pers 34:141–177

Shrout PE, Fleiss JL (1979) Intraclass correlations: Uses in assessing rater reliability. Psychol Bull 2:420–428

Sih A (2011) Behavioral Syndromes: A behavioral ecologist's view on the evolutionary and ecological implications of animal personality. In: Weiss A, King JE, Murray L (eds) Personality and temperament in nonhuman primates. Springer, New York

Stern W (1911) Die differentielle Psychologie in ihren methodischen Grundlagen (2. Auflage). [Differential Psychology in its methodological foundations (2nd ed)]. Barth, Leipzig

Stevenson-Hinde J, Zunz M (1978) Subjective assessment of individual rhesus monkeys. Primates 19:473–482

Stevenson-Hinde J, Stillwell-Barnes R, Zunz M (1980) Individual differences in young rhesus monkeys: consistency and change. Primates 21:498–509

Suomi SJ, Novak MA, Well A (1996) Aging in rhesus monkeys: Different windows on behavioral continuity and change. Dev Psychol 32:1116–1128

Tett RP, Guterman HA (2000) Situation trait relevance, trait expression, and cross-situational consistency: Testing a principle of trait activation. J Res Pers 34:397–423

Tooby J, Cosmides L (1989) Adaptation versus phylogeny: The role of animal psychology in the study of human behavior. Int J Comp Psychol 2:175–188

Uher J (2008a) Comparative personality research: Methodological approaches (Target article). Eur J Pers 22:427–455

Uher J (2008b) Three methodological core issues of comparative personality research. European Eur J Pers 22:475–496
Uher J (2011) Individual behavioral phenotypes: An integrative meta-theoretical framework. Why 'behavioral syndromes' are not analogues of 'personality'. Dev Psychobiol, published online Mar 22, 2011, doi:10.1002/dev.20544
Uher J, Asendorpf JB (2008) Personality assessment in the Great Apes: Comparing ecologically valid behavior measures, behavior ratings, and adjective ratings. J Res Pers 42:821–838
Uher J, Asendorpf JB, Call J (2008) Personality in the behaviour of great apes: Temporal stability, cross-situational consistency, and coherence in response. Anim Behav 75:99–112
van de Vijver FJR, Poortinga YH (2002) Structural equivalence in multilevel research. J Cross Cult Psychol 141:141–156
Vazire S, Gosling SD, Dickey AS et al. (2007) Measuring personality in nonhuman animals. In: Robins RW, Fraley RC, Krueger RF (eds) Handbook of research methods in personality psychology. Guilford Press, New York
Weiss A, Adams MJ (2008) Species of nonhuman personality assessment. Eur J Pers 22:472–474
Weiss A, King JE, Hopkins WD (2007) A cross-setting study of chimpanzee (*Pan troglodytes*) personality structure and development: Zoological parks and Yerkes National Primate Research Center. Am J Primatol 69:1264–1277
Weiss A, King JE, Perkins L (2006) Personality and subjective well-being in orangutans (*Pongo pygmaeus* and *Pongo abelii*). J Pers Soc Psychol 90:501–511
Yerkes RM (1939) The life history and personality of the chimpanzee. Am Nat 73:97–112